

# 1 Determining the Aggregation Function $A$

At first glance, the ranking of the variables  $\mathcal{V}$  seems entirely subjective. It's not hard to buy that this aggregation function  $A$  depends on the given variables, but the manner in which each variable “positively impacts student performance” is very dependent on what it actually *means*. Unfortunately for us, the semantics of the data depend on world knowledge tied into the description, and there's no intrinsic reason to prefer any given variables over any other.

Nonetheless, we might be able to exploit internal consistencies in the data to absolve ourselves of needing to manually assign “goodness” values to every variable. The idea is as follows: even without knowing anything about what variables do, we might be able to group them by how similar they are – how often they ‘stand or fall together’.

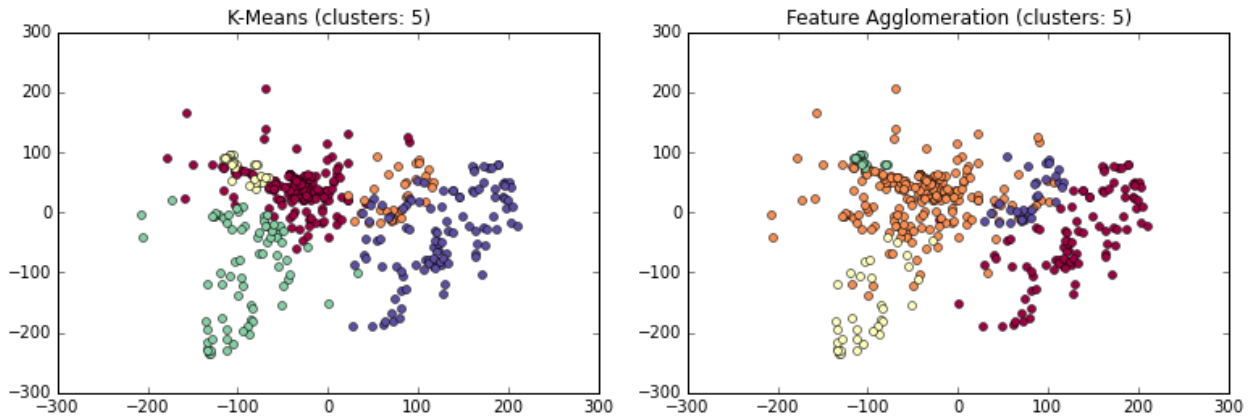
For a given random variable  $X : \mathcal{X} = \{x_i\} \rightarrow \mathbb{R}$ , we have the usual definition of its associated entropy:

$$H(X) = - \sum_{x_i \in \mathcal{X}} P(x_i) \log_2(P(x_i)), \quad (1)$$

## 1.1 Clustering

Clustering, too seems to be an impossible problem, so we

## 1.2 Cluster Separation



## 1.3 Models to Assign Weights